Mammalian mismatch repair: error-free or error-prone?

Peña-Diaz, Javier ; Jiricny, Josef

Abstract: A considerable surge of interest in the mismatch repair (MMR) system has been brought about by the discovery of a link between Lynch syndrome, an inherited predisposition to cancer of the colon and other organs, and malfunction of this key DNA metabolic pathway. This review focuses on recent advances in our understanding of the molecular mechanisms of canonical MMR, which improves replication fidelity by removing misincorporated nucleotides from the nascent DNA strand. We also discuss the involvement of MMR proteins in two other processes: trinucleotide repeat expansion and antibody maturation, in which MMR proteins are required for mutagenesis rather than for its prevention.

DOI: https://doi.org/10.1016/j.tibs.2012.03.001

Mammalian mismatch repair: error-free or error-prone?

Javier Peña-Diaz and Josef Jiricny*

Institute of Molecular Cancer Research
of the University of Zurich and the ETH Zurich

Winterthurerstrasse 190
8057 Zurich
Switzerland

Abstract: 138 words

***Correspondence to: jiricny@imcr.uzh.ch**

Abstract

The discovery of a link between the Lynch Syndrome, an inherited predisposition to cancer of the colon and other organs, and malfunction of the mismatch repair (MMR) system has brought about a considerable surge of interest in this key pathway of DNA metabolism. This review focuses on recent advances in our understanding of the molecular mechanisms of canonical MMR, which improves replication fidelity by removing misincorporated nucleotides from the nascent DNA strand. We also discuss the involvement of MMR proteins in two other processes: trinucleotide repeat expansion and antibody maturation, in which MMR proteins are required for mutagenesis rather than for its prevention.

Introduction

Replication errors represent a considerable threat to genomic integrity. Failure to repair base-base mismatches and insertion/deletion loops (IDLs) arising during DNA replication increases mutation frequencies by 2-3 orders of magnitude. The finding that mismatch repair (MMR) deficiency in *S. cerevisiae* gave rise to microsatellite instability (MSI) led to the prediction that tumors with MSI might also have defective MMR [1]. This was indeed the case: thanks to the high degree of evolutionary conservation of the MMR process, human homologs of yeast *MMR* genes were rapidly identified and shown to be mutated in Lynch Syndrome families that are predisposed to early-onset cancer of the colon, endometrium, ovary and other organs [2, 3]. Its link to cancer brought MMR into the limelight, which led to the identification and phenotypic characterization of MMR-deficient human tumor cell lines, to the expression of recombinant MMR proteins and the determination of their structures, and eventually to the reconstitution of human MMR from purified constituent proteins.

That MMR defects cause cancer was confirmed in knock-out mouse models, but these experiments also yielded some unexpected findings, which implicated MMR proteins in other pathways of DNA metabolism, such as triplet repeat expansion (TRE), somatic hypermutation (SH) and class switch recombination (CSR). MMR involvement in the latter processes was unexpected - how could a high-fidelity DNA repair pathway participate in region-specific mutagenesis? The latter phenomena are currently attracting a great deal of attention, and we shall therefore discuss them at some length. However, an understanding of these processes requires a thorough knowledge of the molecular transactions that take place during the repair of replication errors. This topic will therefore be discussed first.

Repair of base/base mismatches and IDLs.

The minimal human MMR system was reconstituted from its individual purified components

MutSα or MutSβ, MutLα, EXO1, RFC, PCNA, RPA, polymerase-δ and DNA ligase I several years ago [4, 5] and has been extensively reviewed [6-11]. We shall therefore discuss only the key features of the system and focus on recent work that yielded novel mechanistic insights into this important pathway of DNA metabolism.

The first biochemical studies, carried out with extracts of human or *Drosophila melanogaster* cells [12, 13], used circular DNA substrates carrying a single mismatch (Fig. 1A,B). These experiments showed that covalently-closed circular substrates were largely refractory to MMR, but that a nick in either strand situated up to ~1kb away from the mismatch was sufficient to activate the repair process and direct it to the nicked strand. These experiments also showed that MMR can make use of a nick situated 3' (Fig. 1A) or 5' (Fig. 1B) from the misincorporated nucleotide (red G in Fig. 1) [4]. This finding was puzzling, as the only exonuclease implicated in eukaryotic mismatch repair, EXO1, has an obligate 5'➔3' polarity and its loading at a 3' nick would therefore cause it to degrade DNA away from the mismatch rather than towards it. This apparent mystery was explained when the PMS2 subunit of MutLα was shown to possess a cryptic endonuclease activity, which, once activated, introduces additional single-strand breaks into the pre-nicked strand [14]. Armed with this knowledge, it was possible to broadly classify the basic steps of MMR as licensing, degradation and resynthesis. The licensing step is initiated by MutSα binding to the mismatch, which brings about an exchange of ADP for ATP and converts MutSα into a sliding clamp. MutLα is recruited to form a ternary complex, which then most likely diffuses along the DNA contour [15] until it encounters PCNA loaded at a 3'-terminus of a nearby nick by RFC. On 3' substrates, this process leads to activation of the cryptic endonuclease of MutLα and the generation of additional nicks flanking the mismatch (red arrows in step **a** of Fig. 1A). The degradation step (**b** in Fig. 1A,B) involves loading of EXO1 at the nicks by the activated MutSα/MutLα complex, which generates a single-stranded gap starting at the nick and terminating ~150 nucleotides past the mismatch. The resynthesis step (**c** in Fig. 1A,B) involves PCNA, polymerase-δ and DNA ligase I [6-11].

How MutLα introduces additional breaks selectively into the nicked strand was elucidated in the follow-up study [16], which showed that the MutLα endonuclease is activated by interaction with PCNA. This homotrimeric ring, which has two distinct faces, is loaded on the DNA by the clamp-loader RFC, mostly at boundaries between single- and double-stranded DNA [17], and always with the same face oriented towards the 3'terminus (shown light brown in Fig. 1B). Once loaded, the ring is free to rotate around the longitudinal axis of the helix, but it cannot flip around. It therefore follows that when PCNA binds its interaction partners on DNA, the resulting complexes have a fixed orientation, which will be retained even if they were to translocate a large distance away from the site where they were originally loaded. Because MutLα has only a single endonuclease active site (in the PMS2 subunit), it will cleave only one strand of the duplex per association event. Once bound to PCNA, MutLα orientation is fixed and PMS2 will thus be able to nick only the strand running in the 5'➔3' or 3'➔5' direction, depending on the polarity of its endonuclease and on the face of PCNA to which it is bound. The model shown in Figure 1B supposes that PMS2 cleaves only 5'➔3' phosphodiester linkages. In this constellation, its complex with PCNA would always nick the red (5'➔3') strand, irrespective of its distance from the PCNA loading site, or its rotation around the helix axis.

Surprisingly, the above study [16] showed that RFC can load PCNA with low efficiency also onto DNA lacking pre-existing nicks or gaps (e.g. supercoiled substrates, or molecules containing bubbles or stem-loops). Because PCNA is apparently loaded onto these substrates in either orientation, its association with MutLα gives rise to incisions in either strand. The potential biological significance of this finding is discussed below.

The remaining question is why the additional nicks are generated preferentially in the vicinity of the mismatch? The likely explanation is that the endonuclease introducing these nicks is not MutLα/PCNA, but the complex of PCNA with mismatch-activated MutSα/MutLα, which assembles on heteroduplex DNA ([18] and references therein). Because the number of

MutSα/MutLα complexes will be highest near the mismatch, the likelihood of their colliding with PCNA and cleaving DNA will also be higher in mismatch vicinity (Figure 1C). The ability of the MutSα/MutLα complex to generate nicks in DNA has important implications for MMR *in vivo*. PCNA molecules have been reported to remain on the DNA for some time after replication. Because they were loaded either at the 3' termini of Okazaki fragments or of the leading strand, all processivity clamps on one sister chromatid have the same orientation, which is opposite to those PCNA molecules remaining on the other sister chromatid [19]. Extrapolating from the *in vitro* experiments, when mismatch-activated MutSα/MutLα interacts with one of these PCNA molecules, the resulting complex will cleave the nascent DNA strand, even though the nick or gap where PCNA was originally loaded is no longer there. This process thus endows MMR with the correct strand directionality even after Okazaki fragments have been repaired.

But there might be an alternative way of distinguishing the nascent strand from the template, especially as the number of PCNA molecules remaining on the leading strand allele will be very low. We could show that strand breaks or gaps generated during base excision repair (BER) could be utilized for MMR initiation [20]. Most recently, *S. cerevisiae* polymerase-ε, which catalyzes the synthesis of the leading strand [21], was shown to occasionally incorporate also ribonucleotides into DNA during replication [22]. These are efficiently removed by RNase H2 [23], and the possibility that the single-stranded nicks generated by the latter enzyme are utilized by the MMR system as strand discrimination signals should not be ignored.

In the above scenario, MMR-activated strand degradation is catalyzed exclusively by EXO1. However, the mutator phenotype of *Exo1*$^{-/-}$ mouse cells is milder than that of *Msh2*$^{-/-}$ or *Mlh1*$^{-/-}$ cells [24], and it had therefore been suggested that nucleases other than EXO1 might be involved in mammalian MMR. To date, no such enzyme could be identified. However, it could be shown that 5' nick-directed MMR *in vitro* could occur by an EXO1-independent mechanism, which involves strand displacement mediated by polymerase-δ, together with MutSα, MutLα, RPA, RFC and

PCNA (Figure 1D) [25]. Evidence from mammalian systems supporting the existence of EXO1-independent pathways *in vivo* is currently lacking. Genetic data from *S. cerevisiae* indirectly implicated the 3'➔5' proofreading exonuclease activities of the replicative polymerases δ and ε in postreplicative MMR, but this hypothesis was not substantiated in human cell extracts *in vitro* [4].

Repair of IDLs

Repair of IDLs larger than 2-3 extrahelical nucleotides was assumed to be mechanistically identical to that shown in Figure 1C above, the only exception being that lesion recognition would be mediated by MutSβ rather than MutSα. This supposition has been recently questioned in a biochemical study from the Modrich laboratory, in which complexes between MutSα, MutSβ, MutLα, PCNA and a 200 base-pair heteroduplex containing two extrahelical nucleotides were studied by surface plasmon resonance spectroscopy [18]. While both MutSα and MutSβ formed ternary complexes with MutLα on the heteroduplex, addition of PCNA to a pre-formed MutSα/MutLα/heteroduplex assembly resulted in the formation of a quaternary complex, whereas PCNA addition to the MutSβ/MutLα/heteroduplex assembly led to MutLα displacement. As mutation of the MSH3 PIP box compromised the binding of MutSβ to both MutLα and PCNA, it was suggested that MLH1 and PCNA bind to MSH3 at the same (or partially overlapping) site. If the interaction of MutSβ with MutLα and PCNA were mutually exclusive *in vivo*, it would imply that the molecular mechanisms of repair of base/base mismatches and IDLs might differ. This is not inconceivable. If we were to assume that the repair of large IDLs mechanistically resembles that of base/base mismatches, PCNA would be unable to diffuse from a 3' nick past a large loop to form a complex with the IDL-activated MutSβ/MutLα sliding clamp on the 5' side of the structure (Figure 1E). This suggests that large structures bound by MutSβ that present an impassable barrier to PCNA may not be subject to canonical, EXO1-mediated MMR (see below).

MMR proteins in the metabolism of triplet repeats

Expansions of trinucleotide repeats (TNRs) are the underlying cause of around 20 neuromuscular and neurodegenerative disorders [26]. The best known pathogenic TNR expansions concern the CAG/CTG triplets linked to Huntington's disease (HD) and Myotonic dystrophy (DM), but the most common disorder is the Fragile X syndrome, which is linked to the expansion of a CGG repeat. Although the etiology of these syndromes is complex, it is generally believed that the mechanism leading to expansion involves the formation of stem-loop structures arising in single-stranded DNA during replication, transcription and possibly also chromatin remodeling [27].

Because of the importance of MMR for the maintenance of microsatellite repeat stability, which includes also microsatellites consisting of repeated trinucleotides, its involvement in TNR expansion was anticipated. However, there were grounds for skepticism, based primarily on the finding that MSI in MMR-deficient cells involved most frequently losses of one repeat unit per replication, which suggested that MMR was predominantly correcting strand misalignments (loops) arising through slipping back of the template strand by a single repeat. Because MMR is directed to the nascent strand, the degradation and resynthesis of this strand would restore the original microsatellite length as dictated by the template. That MMR corrects hairpin structures, such as might arise at CAG repeats, with similar efficiency and directionality to replication-associated IDLs was substantiated by Li and coworkers in an *in vitro* assay [28] and similar findings were reported by the Pearson laboratory [29], even though the repair efficiency was highest for small loop sizes. However, genetic experiments suggested that MMR might be "hijacked" by the cellular machinery to cause TNR expansion [30].

The above hypothesis was based on the requirement of MutSβ and MutLα for TNR expansion in a mouse model of HD [26, 31] and these factors were implicated in causing TNR instability also in a yeast model, where GAA/TTC repeats implicated in Friedreich's ataxia were shown to cause gross chromosomal rearrangements and the frequency of these events was reduced in strains disrupted at *MSH2*, *MSH3* and *PMS1*, but not *MSH6* loci [32]. In the latter study, the repeats were shown by 2-

D gels to inhibit replication fork progression and the instability was dependent on repeat orientation, such that repeats with the GAA in the lagging strand were significantly less stable than the same repeats in an inverted orientation.

In an embryonal stem cell model, TNR instability was linked to proliferation and appeared to cease when the cells differentiated [33]. This and numerous other studies implicated DNA replication in the expansion phenomenon and, correspondingly, literature contains numerous schemes proposing how TNRs at the replication fork might lead to large expansions. By analogy to other non-B DNA forms [34], it is unknown whether the proposed structures actually exist *in vivo*, but assuming that TNRs do slow the progress of the replication fork, it is conceivable that the MCM2-7 helicase might run ahead of the fork and liberate a long stretch of single-stranded DNA containing the repeats [35]. If these were to form stable stem-loop structures (in one or both strands) as proposed, replication might bypass them and thus give rise to progeny DNA that is shorter than the template (Figure 2A). Unrepaired, these intermediates would give rise to deletions, while MutSβ-initiated MMR would direct the degradation/resynthesis process to the nascent (shorter) strand and restore the original sequence. The only way a large expansion could arise is if the TNRs were copied – partially or completely – more than once, but this would require dissociation and erroneous reannealing of the nascent and template strands and replication restart. For a perfectly base-paired stretch of DNA to dissociate, it would have to be unwound by a helicase (e.g. during transcription), or cleaved and resected as during the repair of double strand breaks (DSBs). Both alternatives could be considered; the replication fork block at the GAA/TTC repeat could cause a collapse of the fork, which might give rise to a one-ended DSB. In the yeast system, DSBs have indeed been observed [32] and shown to be dependent on functional MMR.

But how and where could MMR be involved? It seems likely that expansions arise from structures (cruciforms, stem-loops/loops) containing full-length TNRs in both strands, which these repetitive sequences have a propensity to form. Assuming that, as mentioned above, RFC were to load PCNA

at these structures [16], then its association with MutSβ and MutLα might result in an endonucleolytic cleavage of either strand (Figure 2B), which would cause collapse of the structure and generate a 3' terminus that could prime repair synthesis (Figure 2C). A subset of the cleaved structures could give rise to heteroduplexes with one strand containing a large insertion (Figure 2D), the repeated processing of which by the MMR proteins and repair polymerase(s) would generate an expanded TNR.

Although the strand break initiating the expansion was postulated to be introduced by the endonuclease activity of MutLα, mouse genetics also implicated 8-oxoguanine DNA glycosylase Ogg1 in TNR expansion [36]. The high metabolic rate and thus the elevated amount of reactive oxygen species in neurons are believed to give rise to oxidized bases in the DNA such as 8-oxoguanine, the removal of which by Ogg1 generates a strand break. Rather than being further processed by the BER machinery, these strand breaks might in rare instances trigger strand displacement or -degradation, which might, in turn, lead to TNR expansion [36]. In these instances, the process might not require the MutLα endonuclease for generation of the strand breaks.

MMR proteins in antibody maturation

Generation of the immense antibody repertoire in vertebrates entails three key steps: VDJ recombination, somatic hypermutation (SHM) and class switch recombination (CSR). The finding that the latter steps require also MMR proteins came as a surprise, given that MMR has hitherto been believed to be an error-free rather than an error-prone pathway. As a number of reviews on the subject are available [37-39], we shall focus solely on the mechanistic aspects of the processes and on the possible involvement of MMR proteins in them.

Upon completion of VDJ recombination, antigen-stimulated B cells undergo SHM, which gives rise to many mutations in the variable regions of the rearranged immunoglobulin light and heavy chain loci. SHM is triggered by induction of activation-induced cytidine deaminase (AID), which converts deoxycytidines at certain sequence motifs to deoxyuridines. The finding that AID targets

preferentially single-stranded DNA and that both SHM and CSR require active transcription suggests that the C to U deaminations take place in or behind the transcription bubble. Reannealing of the two strands behind the moving transcription bubble gives rise to U/G mispairs. Because these lesions did not arise during replication, they should be addressed exclusively by BER, which should replace the dUMP residues with dCMP and thus restore the C/G base pairs. However, BER in B cells appears to be inefficient [40]. Should some uracil residues remain in the DNA until the next round of replication, they would give rise to C/G➔T/A transition mutations. Should the uracils be removed, but the remaining apyrimidinic (AP) site remain unrepaired, replication/repair synthesis across the abasic site would give rise primarily to C/G to G/C transversions, but also to other types of mutations, depending on which polymerase catalyzed the AP-site bypass. Importantly, all above-mentioned mutations would have arisen at the site of the original, AID-catalyzed deamination. But about 50% of mutations linked to SHM and CSR arise at T/A base pairs, i.e. at sites not deaminated by AID, and it is this subset of mutations that has been genetically linked to MMR (primarily to MSH2, MSH6 and EXO1) and to the translesion DNA polymerases pol-ζ and pol-η [37-39]. It has been proposed that processing of the AID-generated U/G mispairs by MMR would give rise to long repair tracts that might be filled-in by error-prone polymerases to generate the observed mutations. This hypothesis was incompatible with our understanding of MMR; as mentioned above, strand degradation during MMR absolutely requires a pre-existing nick, either for loading of EXO1, or as a site where polymerase-δ catalyzed strand displacement might commence. U/G mispairs arising during transcription should not have nicks in the vicinity (Figure 3Aa). Thus, although U/G mispairs are efficiently recognized by MutSα [41, 42], this binding is futile, due to the lack of an EXO1 loading site. However, given that AID acts processively, it might generate several U/G mispairs in close proximity to one another. We therefore considered the possibility that MMR and BER might compete for these substrates. Specifically, that MutSα/MutLα activated by binding to a U/G mismatch might "hijack" a cleaved abasic site generated by the sequential action of uracil DNA glycosylase (UDG) and AP-endonuclease APE1 during the processing of a nearby U/G

mismatch that was being processed by BER (Figure 3Ac). This was indeed the case *in vitro*: when BER was inhibited, so was MMR [20]. This study clearly demonstrated that BER intermediates can be utilized as EXO1 entry points during MMR.

Should the MMR generated single-stranded DNA gap contain other partially-processed sites of AID action (Figure 3B), repair synthesis might introduce or fix mutations at these positions (Figure 3C). In DNA regions containing two or more U/G mispairs in close proximity, in which the uracil residues are situated in opposite strands of the DNA duplex, the collision of the EXO1-catalyzed strand degradation tract initiating at a cleaved AP-site with a similar site in the other strand would give rise to a double-strand break (Fig. 3D), which is believed to be the initiating signal for CSR [43].

If MMR indeed uses BER intermediates as initiation sites [20], how and why does the normally error-free repair of the exonuclease-generated gaps suddenly become error-prone, such that it might lead to extensive mutagenesis at the immunoglobulin loci? The answer to this question may be forthcoming: MMR protein concentration in cells has been reported to increase during S-phase [44] and they appear to be recruited to replicating DNA in a PCNA-dependent manner, at least *in vitro* [45]. MMR proteins have also been described to be recruited to chromatin in response to DNA damage [46-48], whereupon PCNA appears to be monoubiquitylated to a small extent [47, 48]. The latter post-translational modification has been associated with translesion DNA synthesis [49], whereby blocked high-fidelity replicative polymerases are displaced by less-processive, often error-prone enzymes that are able to bypasss the damage and thus enable replication to resume. These translesion polymerases associate preferentially with monoubiquitylated PCNA. Mouse genetic experiments clearly demonstrated that SHM requires, in addition to MSH2 [50], MSH6 [51], EXO1 [52] and polymerase-η [53, 54] also mono-ubiquitylated PCNA [55]. As SHM mutations at T/A base pairs are reduced to a similar extent in knock-in mice expressing a non-ubiquitylatable PCNA K164R variant, and in *Msh2*$^{-/-}$ or *Msh6*$^{-/-}$ animals [55], the origin of the mutations appears to be monoubiquitylated PCNA-mediated recruitment of error-prone DNA polymerases to MMR-

generated gaps (see for [56] recent review). Why the processing of AID-generated U/G mispair might trigger PCNA ubiquitylation and error-prone DNA synthesis, rather than deploy error-free replicative polymerases is currently the subject of intense study in several laboratories.

Conclusions

The detailed understanding of MMR biochemistry has enabled us not only to appreciate the sophistication of the system in processing biosynthetic errors, but also to use this knowledge to try and predict how MMR proteins might be involved in other processes of DNA metabolism such as TNR expansion, SHM and CSR. These areas are highly relevant to human health and deserve much closer examination. Analysis of the interactome of MMR proteins [57] indicated that these polypeptides associate with partners that play key roles in other DNA damage-processing pathways, such as the repair of interstrand cross-links [57-61]. The role of MMR in this important metabolic pathway is not understood and deserves attention. Another emerging topic is the interplay of MMR with chromatin. There are indications that MMR proteins interfere with CAF-1-dependent nucleosome loading [62] and possibly also with nucleosome repositioning, depending on sequence context [63, 64]. We also identified a direct interaction between CAF-1 and MSH6, but its significance is yet to be elucidated [65]. The role of MMR proteins in DNA damage signaling [66] and in the processing of modified bases [67, 68] has not been discussed here, but this topic remains enigmatic and much remains to be discovered. As the MMR field is attracting much attention at the present time, it is likely that further important discoveries are just around the corner.

Acknowledgements

The authors thank Stephanie Bregenhorn for critical reading of the manuscript. The financial support of the 7th Framework Program of the European Community (grant nr. 223545 “MM2M”) and the Swiss National Science Foundation (grant nr. 310030B-133123) is also gratefully acknowledged.

References

1. Jiricny, J. (1994) Colon cancer and DNA repair: have mismatches met their match? *Trends in genetics : TIG*, 10, 164-168
2. Peltomaki, P.*, et al.* (2004) Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database. *Dis Markers*, 20, 269-276
3. Bonadona, V.*, et al.* (2011) Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *Jama*, 305, 2304-2310
4. Constantin, N.*, et al.* (2005) Human mismatch repair: reconstitution of a nick-directed bidirectional reaction. *The Journal of biological chemistry*, 280, 39752-39761
5. Zhang, Y.*, et al.* (2005) Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell*, 122, 693-705
6. Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*, 7, 335-346
7. Li, G.M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res*, 18, 85-98
8. Iyer, R.R.*, et al.* (2006) DNA mismatch repair: functions and mechanisms. *Chem Rev*, 106, 302-323
9. Kunkel, T.A.*, et al.* (2005) DNA mismatch repair. *Annu Rev Biochem*, 74, 681-710
10. Hsieh, P.*, et al.* (2008) DNA mismatch repair: molecular mechanism, cancer, and ageing. *Mech Ageing Dev*, 129, 391-407
11. Modrich, P. (2006) Mechanisms in eukaryotic mismatch repair. *J Biol Chem*, 281, 30305-30309
12. Thomas, D.C.*, et al.* (1991) Heteroduplex repair in extracts of human HeLa cells. *The Journal of biological chemistry*, 266, 3744-3751
13. Holmes, J., Jr.*, et al.* (1990) Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 5837-5841
14. Kadyrov, F.A.*, et al.* (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*, 126, 297-308
15. Pluciennik, A.*, et al.* (2007) Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 12709-12713
16. Pluciennik, A.*, et al.* (2010) PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. *Proc Natl Acad Sci U S A*, 107, 16066-16071
17. Bowman, G.D.*, et al.* (2004) Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature*, 429, 724-730
18. Iyer, R.R.*, et al.* (2010) MutLalpha and proliferating cell nuclear antigen share binding sites on MutSbeta. *J Biol Chem*, 285, 11730-11739
19. Shibahara, K.*, et al.* (1999) Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell*, 96, 575-585
20. Schanz, S.*, et al.* (2009) Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. *Proc Natl Acad Sci U S A*, 106, 5593-5598
21. Pursell, Z.F.*, et al.* (2007) Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*, 317, 127-130
22. Nick McElhinny, S.A.*, et al.* (2010) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 4949-4954
23. Nick McElhinny, S.A.*, et al.* (2010) Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol*, 6, 774-781
24. Wei, K.*, et al.* (2003) Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. *Genes Dev*, 17, 603-614

25. Kadyrov, F.A.*, et al.* (2009) A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proc Natl Acad Sci U S A*, 106, 8495-8500
26. McMurray, C.T. (2010) Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet*, 11, 786-799
27. Lopez Castel, A.*, et al.* (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature reviews. Molecular cell biology*, 11, 165-170
28. Tian, L.*, et al.* (2009) Mismatch recognition protein MutSbeta does not hijack (CAG)n hairpin repair in vitro. *J Biol Chem*, 284, 20452-20456
29. Panigrahi, G.B.*, et al.* (2010) Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. *Proc Natl Acad Sci U S A*, 107, 12593-12598
30. McMurray, C.T. (2008) Hijacking of the mismatch repair system to cause CAG expansion and cell death in neurodegenerative disease. *DNA Repair (Amst)*, 7, 1121-1134
31. Dragileva, E.*, et al.* (2009) Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol Dis*, 33, 37-47
32. Kim, H.M.*, et al.* (2008) Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *Embo J*, 27, 2896-2906
33. Seriola, A.*, et al.* (2011) Huntington's and myotonic dystrophy hESCs: down-regulated trinucleotide repeat instability and mismatch repair machinery expression upon differentiation. *Human molecular genetics*, 20, 176-185
34. Wells, R.D. (2007) Non-B DNA conformations, mutagenesis and disease. *Trends Biochem Sci*, 32, 271-278
35. Byun, T.S.*, et al.* (2005) Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev*, 19, 1040-1052
36. Kovtun, I.V.*, et al.* (2007) Crosstalk of DNA glycosylases with pathways other than base excision repair. *DNA Repair (Amst)*, 6, 517-529
37. Peled, J.U.*, et al.* (2008) The biochemistry of somatic hypermutation. *Annu Rev Immunol*, 26, 481-511
38. Di Noia, J.M.*, et al.* (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem*, 76, 1-22
39. Longerich, S.*, et al.* (2006) AID in somatic hypermutation and class switch recombination. *Current opinion in immunology*, 18, 164-174
40. Poltoratsky, V.*, et al.* (2007) Down-regulation of DNA polymerase beta accompanies somatic hypermutation in human BL2 cell lines. *DNA repair*, 6, 244-253
41. Hughes, M.J.*, et al.* (1992) The purification of a human mismatch-binding protein and identification of its associated ATPase and helicase activities. *The Journal of biological chemistry*, 267, 23876-23882
42. Larson, E.D.*, et al.* (2005) MutSalpha binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions. *Current biology : CB*, 15, 470-474
43. Stavnezer, J. (2011) Complex regulation and function of activation-induced cytidine deaminase. *Trends in immunology*, 32, 194-201
44. Edelbrock, M.A.*, et al.* (2009) DNA mismatch repair efficiency and fidelity are elevated during DNA synthesis in human cells. *Mutat Res*, 662, 59-66
45. Masih, P.J.*, et al.* (2008) Mismatch Repair proteins are recruited to replicating DNA through interaction with Proliferating Cell Nuclear Antigen (PCNA). *Nucleic acids research*, 36, 67-75
46. Hong, Z.*, et al.* (2008) Recruitment of mismatch repair proteins to the site of DNA damage in human cells. *J Cell Sci*, 121, 3146-3154
47. Mastrocola, A.S.*, et al.* (2010) Nuclear reorganization of DNA mismatch repair proteins in response to DNA damage. *DNA Repair (Amst)*, 9, 120-133
48. Schroering, A.G.*, et al.* (2008) Rapid induction of chromatin-associated DNA mismatch repair proteins after MNNG treatment. *DNA Repair (Amst)*, 7, 951-969

49. Lehmann, A.R. (2011) Ubiquitin-family modifications in the replication of DNA damage. *FEBS Lett*,
50. Roa, S.*, et al.* (2010) MSH2/MSH6 complex promotes error-free repair of AID-induced dU:G mispairs as well as error-prone hypermutation of A:T sites. *PloS one*, 5, e11182
51. Wiesendanger, M.*, et al.* (2000) Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *The Journal of experimental medicine*, 191, 579-584
52. Bardwell, P.D.*, et al.* (2004) Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat Immunol*, 5, 224-229
53. Delbos, F.*, et al.* (2007) DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *The Journal of experimental medicine*, 204, 17-23
54. Delbos, F.*, et al.* (2005) Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. *The Journal of experimental medicine*, 201, 1191-1196
55. Langerak, P.*, et al.* (2009) Somatic hypermutation of immunoglobulin genes: lessons from proliferating cell nuclear antigenK164R mutant mice. *Philos Trans R Soc Lond B Biol Sci*, 364, 621-629
56. Chahwan, R.*, et al.* (2011) Mismatch-mediated error prone repair at the immunoglobulin genes. *Biomed Pharmacother*,
57. Cannavo, E.*, et al.* (2007) Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2. *J Biol Chem*, 282, 2976-2986
58. Kratz, K.*, et al.* (2010) Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell*, 142, 77-88
59. Liu, T.*, et al.* (2010) FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science*, 329, 693-696
60. MacKay, C.*, et al.* (2010) Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. *Cell*, 142, 65-76
61. Smogorzewska, A.*, et al.* (2010) A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Molecular cell*, 39, 36-47
62. Kadyrova, L.Y.*, et al.* (2011) CAF-I-dependent control of degradation of the discontinuous strands during mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 2753-2758
63. Javaid, S.*, et al.* (2009) Nucleosome remodeling by hMSH2-hMSH6. *Mol Cell*, 36, 1086-1094
64. Li, F.*, et al.* (2009) Evidence that nucleosomes inhibit mismatch repair in eukaryotic cells. *J Biol Chem*, 284, 33056-33061
65. Schöpf, B.*, et al.* (2012) The interplay between mismatch repair and chromatin assembly. *Proc Natl Acad Sci U S A*, in press
66. Stojic, L.*, et al.* (2004) Mismatch repair and DNA damage signalling. *DNA repair*, 3, 1091-1101
67. York, S.J.*, et al.* (2006) Mismatch repair-dependent iterative excision at irreparable O6-methylguanine lesions in human nuclear extracts. *J Biol Chem*, 281, 22674-22683
68. Zhao, J.*, et al.* (2009) Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. *Nucleic Acids Res*, 37, 4420-4429

Legends to Figures

Figure 1: *Schematic representation of the key MMR steps.* In cell-free systems, the circular heteroduplex substrates require a nick or a gap in addition to the mispair. A nick either 3' (**A**), or 5' (**B**) from the mispaired G in the outer strand will result in a G/T to A/T repair (as shown). A nick in the inner strand would result in a G/T to G/C repair. **A**, MutSα (MSH2/MSH6 heterodimer) binds the mismatch, recruits MutLα (MLH1/PMS2 heterodimer) and translocates along the DNA contour until it encounters PCNA bound at the 3' terminus of the nick. (a) The ternary complex introduces additional breaks into the nicked strand (red arrows), where EXO1 is loaded. EXO1 generates a long single-stranded gap (b), which is filled-in by the PCNA/polymerase-δ complex (c) and the remaining nick is then sealed. In this scenario, the strand degradation reaction initiates at one of the breaks generated *de novo* by the mismatch-activated MutSα/MutLα/PCNA complex. **B**, When the nick is situated 5' from the G, mismatch-activated MutSα/MutLα can load EXO1 directly at the nick. In this scenario, strand degradation (b) will terminate ~150 nucleotides past the mismatch. The gap is filled-in by the PCNA/polymerase-δ complex (c) as in A. **C**, MutSα/MutLα PCNA loaded at the 3' terminus of a nick recruits MutLα. Because PCNA is loaded in a directional manner (with the light brown face towards the 3' terminus of the nicked strand as shown here), and because its interaction site with MutLα is also invariable, the resulting complex will have a given orientation even if it were to translocate along the DNA contour. As the PMS2 subunit of MutLα can cleave only one type of phosphodiester linkage (shown here as 5' to 3'), it will cleave always the same strand, irrespective of its distance from the PCNA loading site, or its rotational orientation on the DNA helix. **D**, When the terminus of the nascent DNA strand is situated on the 3' side of the misincorporated nucleotide (G), such as in the leading strand that is being synthesized by pol-ε as shown, MutLα has to introduce additional breaks into the nascent strand, some of which have to be on the 5' side of the G. Because the mismatch-activated MutSα sliding clamp can diffuse away in either direction (green arrow) from the mispair (probably together with MutLα as shown here) and because the mispair (or a small IDL) presents no barrier to PCNA (or to PCNA/MutLα complex) diffusion, the PMS2 endonuclease will be activated anywhere where the quaternary complex forms. On a leading strand, PCNA could not diffuse beyond the 3' terminus of the leading strand. On a substrate pictured in A above, RFC bound at the 5' terminus of the nick (as in E below) may block translation of PCNA in the 5' to 3' direction. **E**, EXO1-independent MMR. When the nick is positioned 5' from the mispaired G, repair can occur simply through a MutSα-activated and polymerase-δ catalyzed strand-displacement reaction, followed by a FEN1-catalyzed flap removal and nick-sealing by DNA ligase I. **F**, Communication between MutSβ (MSH2/MSH3 heterodimer), MutLα and PCNA on substrates containing large loops may be hindered. The molecular mechanism of the repair process may thus differ from that of mispairs and small IDLs.

Figure 2: *Simplified insights into the metabolism of TNRs.* **A**, Should TNRs delay the progression of the replicating polymerases, the MCM2-7 helicase might get ahead, leaving behind long single-stranded regions, which might fold into secondary structures (red stem-loops). A by-pass of these structures would give rise to nascent DNA (green) that is substantially shorter than the template. Because canonical MMR is directed to the nascent strand, degradation of this strand (green) and repair synthesis would restore the template strand sequence, rather than lead to TNR expansion. **B**, Cleavage of a cruciform structure arising in a TNR sequence (red) would result in its collapse. In a subset of events, the newly-generated 3' terminus could anneal to the complementary strand (**C)** and prime DNA synthesis (**D**) that would lead to a substantial expansion of the repeat sequence in one strand.

Figure 3: *Possible DNA substrates arising during SHM and CSR.* **A**, Following the passage of AID, most likely in complex with the transcription machinery of RNA Polymerase II (RNAPII), and reannealing of the two strands of the transcription bubble, the U/G mispairs (a) can be addressed by uracil DNA glycosylase (UDG) to leave behind an apyrimidinic site (b, red bar). Cleavage of this site by AP-endonuclease (APE1) would leave a strand break (c), which is normally processed further by polymerase-β and DNA ligase III/XRCC1 (Lig III) to regenerate the C/G pair. This process would be error-free. **B**, If the U/G-activated MMR were to load EXO1 at the site of the cleaved AP-site (c), the single-stranded gap might contain a number of different modifications (uracils or AP-sites). **C**, Filling-in of the single-stranded gap may not be possible by high-fidelity polymerases. Their arrest might trigger PCNA ubiquitylation, which would recruit translesion polymerases. These might in turn generate mutations, not only at the original sites of AID action, but also elsewhere in the resynthesized patch. **D**, In cases where the EXO1-catayzed degradation encountered a cleaved AP-site in the opposite strand, a double-strand break (red lightning) would arise, which might trigger CSR.

Boxes

IDLs - insertion/deletion loops; loops of extrahelical nucleotides arising during replication through slippage of the temple and primer strands. Uncorrected, theee loops would give rise to either insertions or deletions in the progeny DNA.

MSI – microsatellites are repetitive sequences consisting mostly of runs of mono-, di-, or trinucleotides, which are present in many thousands in higher eukaryotic genomes. Most polymerases have difficulty in duplicating these DNA stretches with high fidelity and tend to slip, giving rise to IDLs. These are corrected with high efficiency by MMR, but in its absence give rise to insertions or deletions and thus to fluctuations in microsatellite repeat numbers.

MutSα - Heterodimer of MSH2 and MSH6, eukaryotic homologs of the bacterial mismatch-binding protein MutS. This factor binds base-base mismatches and IDLs of 1-3 nucleotides (depending on sequence context), whereupon it undergoes a conformational change, which converts it into a sliding clamp on the DNA. This alteration is triggered by mismatch binding and is accompanied by an exchange of bound ADP for ATP.

MutSβ - Heterodimer of MSH2 and MSH3, eukaryotic homologs of the bacterial mismatch-binding protein MutS. This factor binds IDLs of 1to approximately12 nucleotides (depending on sequence), whereupon it undergoes a conformational change, which converts it into a sliding clamp on the DNA. This alteration is triggered by mismatch binding and is accompanied by an exchange of bound ADP for ATP. It may bind also larger IDLs or stem-loops structures, but in this case, it often fails to undergo the ATP-driven conformational change.

MutLα - Heterodimer of MLH1 and PMS2, eukaryotic homologs of the bacterial mismatch-repair protein MutL.

EXO1 – Exonuclease 1, a 5’ to 3’ exonuclease that participates in MMR, as well as in homologous recombination and resection of double-strand breaks.

PCNA – Homotrimeric protein that forms a ring around DNA and that associates with numerous polypeptides *via* a so-called PCNA-interacting peptide (PIP) motif QxxLxxFF. Its best known function is that of a processivity clamp for replicative polymerases.

RFC – Replication Factor C; a heteropentameric protein otherwise known as “clamp-loader”. RFC uses the energy of ATP hydrolysis to load PCNA onto DNA at free 3’ termini.

RPA – Replication Protein A; a heterotrimeric single-strand binding protein.

Okazaki fragments – During replication, the fork movement dissociates the two strands of the DNA duplex. Due to the antiparallel orientation of the two strands, and to the fact that DNA synthesis occurs only in the 5’ to 3 direction, only one strand, defined as the leading strand, can be copied by the replicating polymerase in a continuous manner. DNA synthesis on the lagging strand is disconituous. Immediately behind the replication fork, the nascent lagging strand consists of a series of 200-300 nucleotide long fragments named after their discoverer. These are eventually joined to yield a continuous lagging strand.

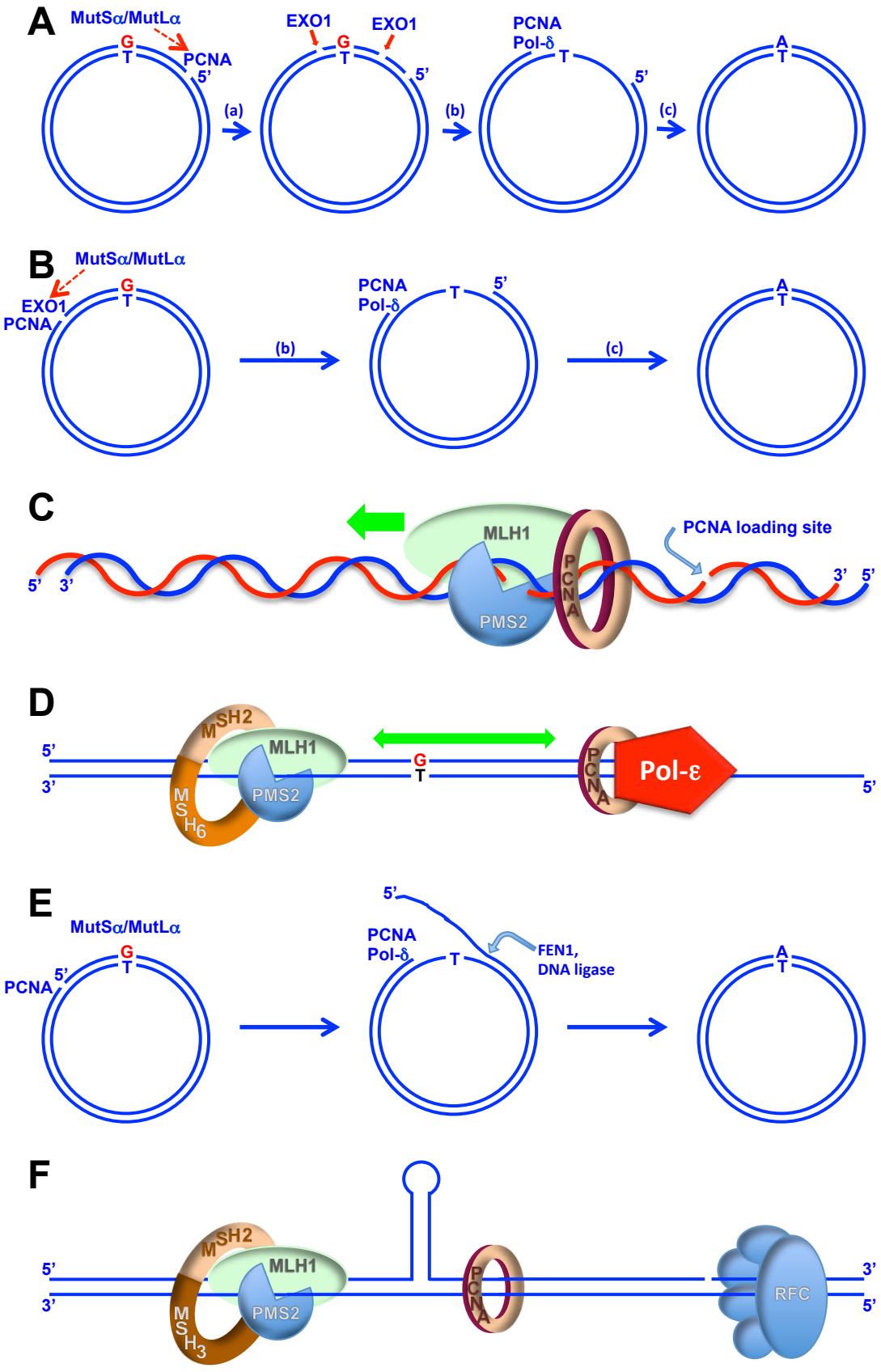
BER – base excision repair; this process is specific for the removal of aberrant bases from DNA and involves the replacement of 1-6 nucleotides. The aberrant base is removed by a specific DNA glycosylase. The resulting abasic (apyrimidinic or apurinic, AP) site is subsequently cleaved by AP-endonuclease, APE1. In “short-patch BER”, a single nucleotide is incorporated by polymerase-β, which concurrently removes the baseless sugar-phosphate. The remaining nick is sealed by DNA ligase III/XRCC1. In “long-patch BER”, the polymerase displaces 2-6 nucleotides. The resulting flap is cleaved off by FEN1 and the remaining nick is sealed by DNA ligase I.

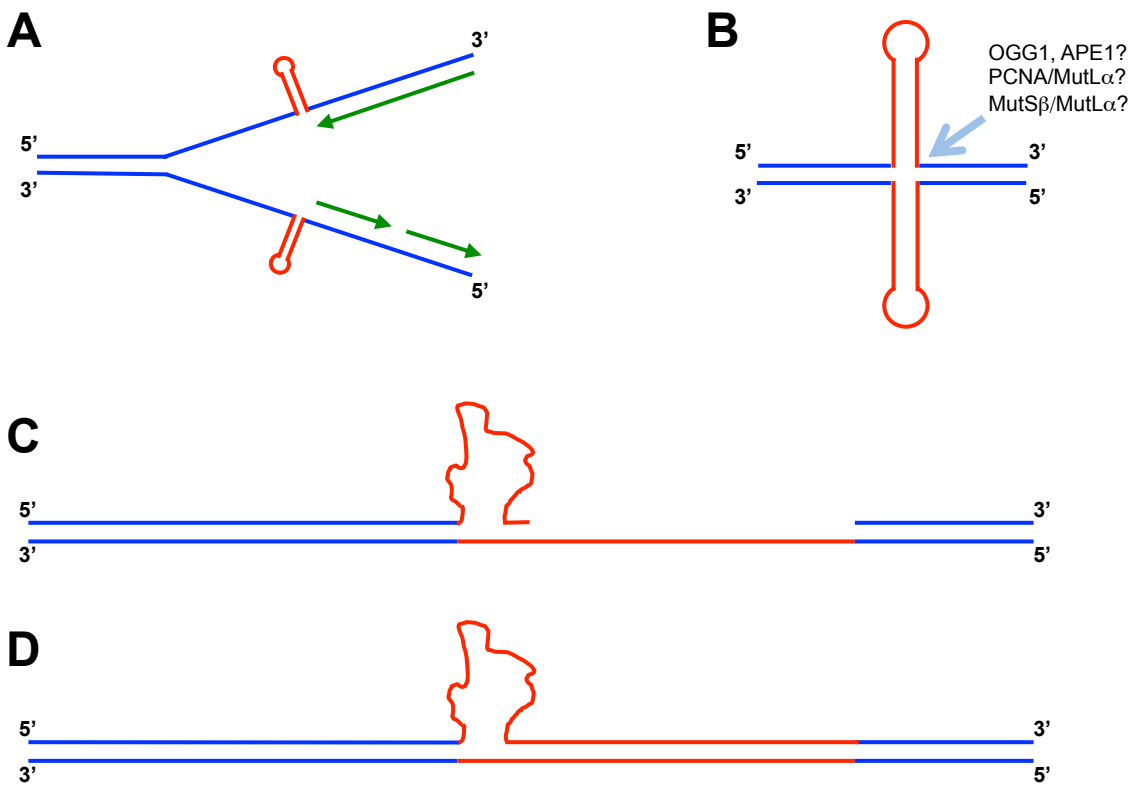
VDJ recombination – A recombination process taking place in pre-B cells in the bone marrow between the numerous variable (V), diversity (D) and join (J) domains of the immunoglobulin genes, which results in the expression of a variety of IgM antibodies. This process is not antigen-dependent.

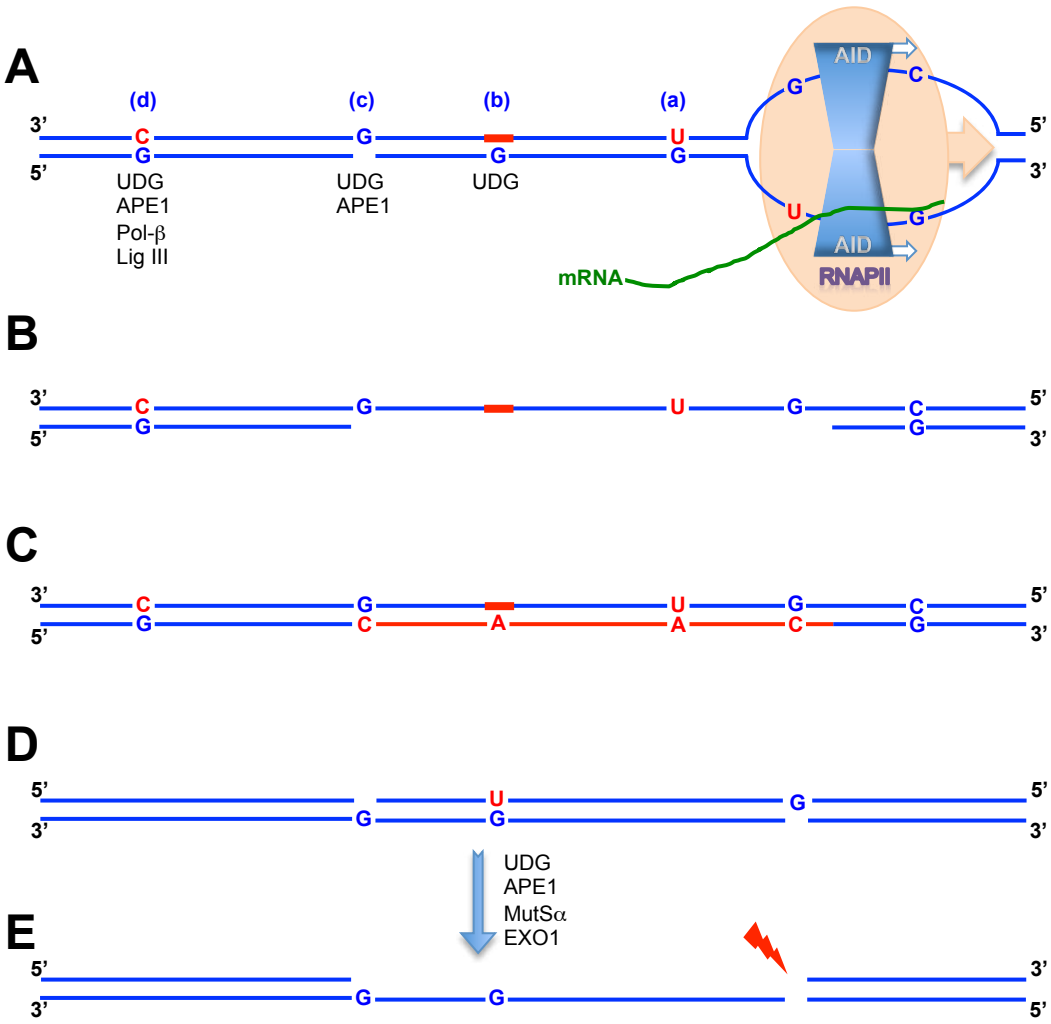
SHM – somatic hypermutation; a mutagenic process confined largely to the recombined VDJ regions in antigen-stimulated B cells in germinal centers. This random process of mutagenesis is initiated by activation-induced cytidine deaminase (AID), which converts numerous cytidines in these regions to uridines.

CSR – class switch recombination; AID-induced deamination of cytidines in the highly-homologous so-called switch regions of immunoglobulin genes, which triggers recombination events that lead to the change from IgM antibody isotype to IgG, IgA and IgE isotype expression.

UDG – Uracil DNA-glycosylase; an enzyme that specifically removes uracil from DNA, without cleaving the sugar-phosphate backbone. It leaves behind an apyrimidinic (AP) site.


A
MutSα/MutLα
G
T
PCNA
5'
(a)
EXO1
EXO1
G
T
5'
(b)
PCNA
Pol-δ
T
5'
(c)
A
T
B
MutSα/MutLα
EXO1
PCNA
G
T
(b)
PCNA
Pol-δ
T
5'
(c)
A
T
C
MLH1
PCNA loading site
5'
3'
PMS2
PCNA
3'
5'
D
MSH2
MLH1
5'
3'
PMS2
MSH6
G
T
PCNA
Pol-ε
5'
E
MutSα/MutLα
PCNA
5'
G
T
5'
PCNA
Pol-δ
T
FEN1,
DNA ligase
A
T
F
MSH2
MLH1
5'
3'
PMS2
MSH3
PCNA
RFC
3'
5'


A
3'
5'
3'
5'
B
OGG1, APE1?
PCNA/MutLα?
MutSβ/MutLα?
5'
3'
3'
5'
C
5'
3'
3'
5'
D
5'
3'
3'
5'


A
(d)
(c)
(b)
(a)
3'
5'
C
G
G
U
G
UDG
APE1
Pol-β
Lig III
UDG
APE1
UDG
AID
AID
RNAPII
mRNA
B
C
A
D
UDG
APE1
MutSα
EXO1
E